AWS re:Invent

STG355

Archiving media content with Amazon S3 Glacier Deep Archive

Rob Czarnecki
Principal Product Manager
Amazon Web Services, Inc.

Tanuja Korelpra
Sr. Product Manager
Amazon Web Services, Inc.

Heidi Quicksilver
Senior Director Digital Systems and Strategy
Rock & Roll Hall of Fame

AWS re:Invent

Agenda

AWS storage overview

Archiving with Amazon Simple Storage Service (Amazon S3) Glacier Deep Archive

Customer use case: Rock & Roll Hall of Fame

Ingest paths

Best practices and solutions

Questions

AWS storage overview

AWS re:Invent

aws

AWS storage portfolio

OBJECT

Amazon S3

BLOCK

Amazon EBS

FILE

Amazon EFS

Amazon FSx for Windows

Amazon FSx for Lustre

BACKUP

AWS Backup

DATA TRANSFER AND EDGE PROCESSING

AWS Storage Gateway

AWS DataSync

AWS Transfer for SFTP

AWS Snowmobile

AWS Snowball

AWS Snowball Edge

Amazon S3 offers industry-leading scalability, availability, security, and performance



Exabytes stored across many millions of drives

S3 stores trillions of objects around the world

We regularly peak at millions of requests per second

In a single region, we process peaks of over 60 tbps in a day

235+ distributed micro-services in 3 or more AZs per Region

Content is growing rapidly



Media



User data



Files



Connected devices

You HAVE to have an archiving plan

Archiving with Amazon Simple Storage Service (S3)

AWS re:Invent

aws

Your choice of Amazon S3 storage classes

Frequent ← → Infrequent ← → Archive

S3 Standard



Frequent

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB

S3 Intelligent-Tiering



- Data with changing access patterns
- Milliseconds access
- ≥ 3 AZ
- $0.0210 to $0.0125/GB
- Monitoring fee per object
- Min storage duration

S3 Standard-IA



Infrequent

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 One Zone-IA



- Re-creatable, less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier



- Archive data
- Select minutes or hours
- ≥ 3 AZ
- $0.0040/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier Deep Archive



Archive

- Long term archive data
- Select hours
- ≥ 3 AZ
- $0.00099/GB
- Retrieval fee per GB
- Min storage duration
- Min object size

S3 Glacier Deep Archive: Lowest-cost cloud storage

No tape to manage

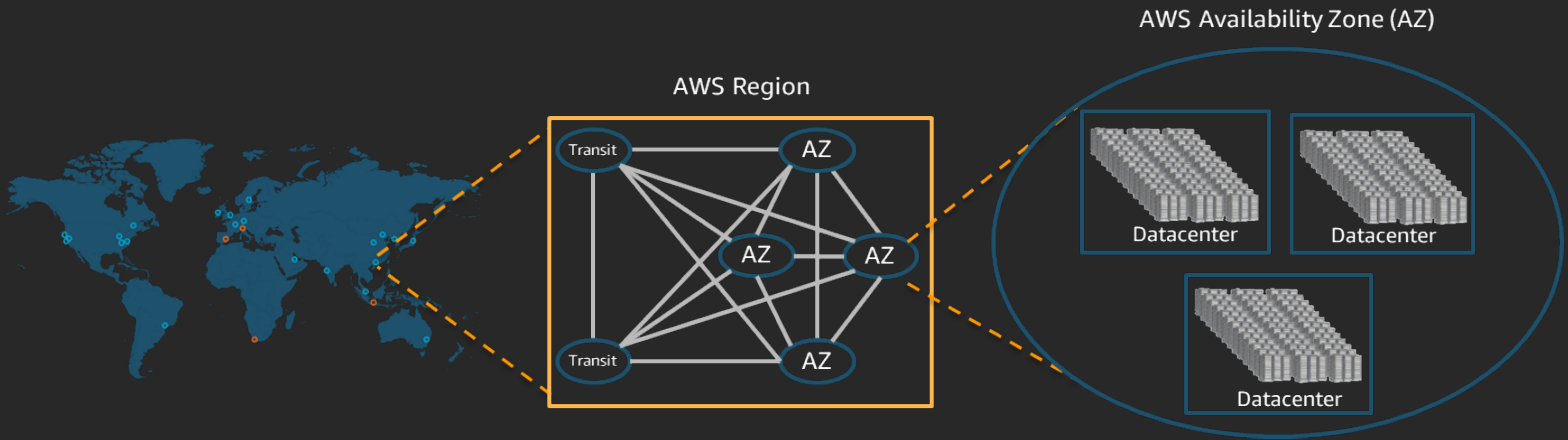
Designed for 99.999999999% durability

Recover data in hours vs. days/weeks

$0.00099 per GB-month

Less than 1/4 the cost of S3 Glacier

Scale globally with resilience in every region

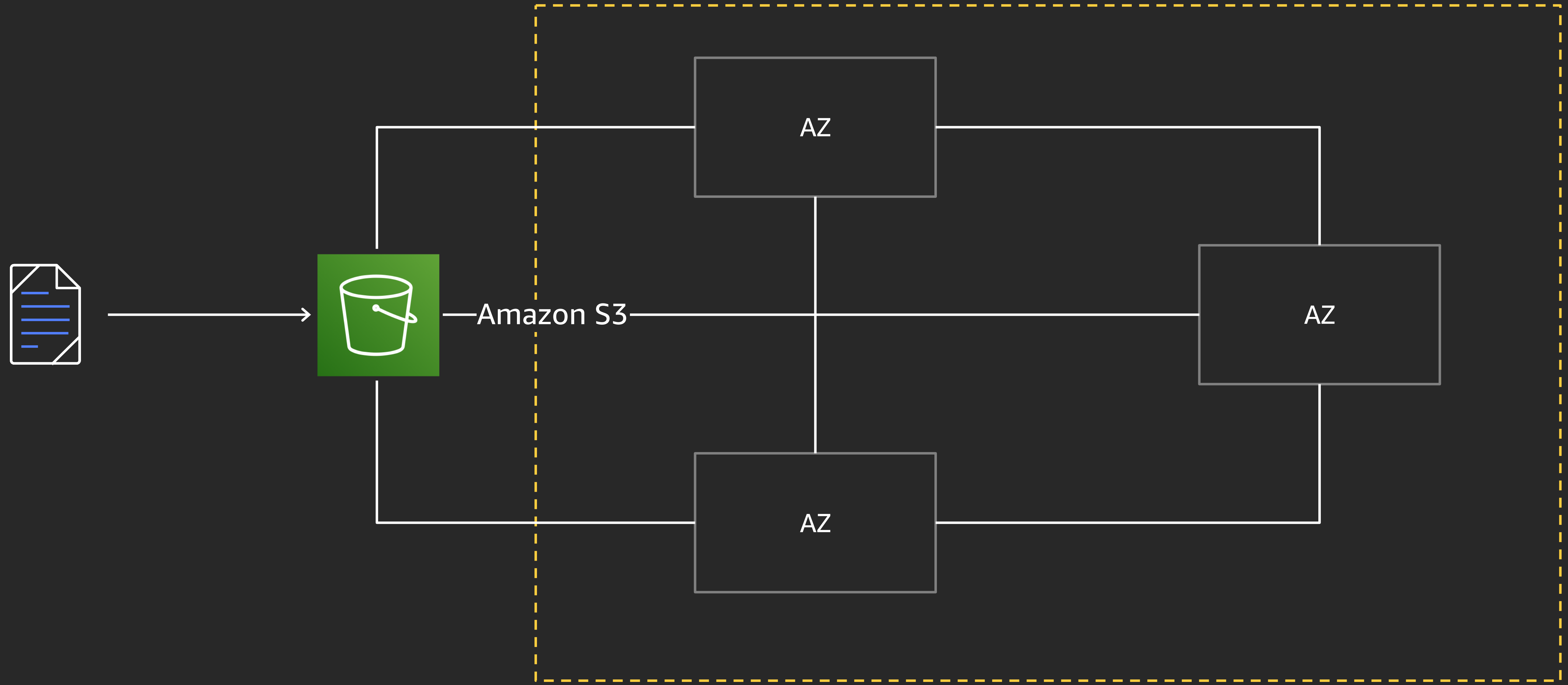


AZs separated by miles

Redundant, dedicated network interconnect

Each AZ has independent power infrastructure

Benefits of 3 independent AZs in each AWS Region

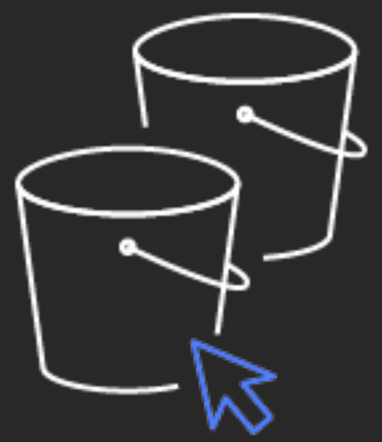


99.9% Availability SLA

Designed for
99.999999999% durability

Ingest paths

Moving data into S3 Glacier Deep Archive



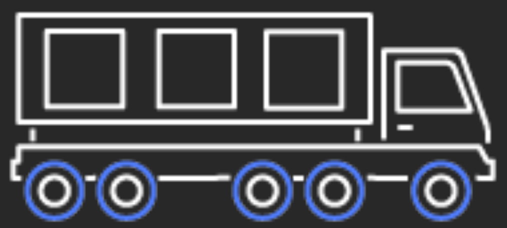
PUT, COPY



Replication



Lifecycle



AWS data transfer services



APN

Archival storage through Amazon S3 API

Write to any S3 storage class directly via the Amazon S3 API

Direct PUT, replicate, or lifecycle via Amazon S3 API

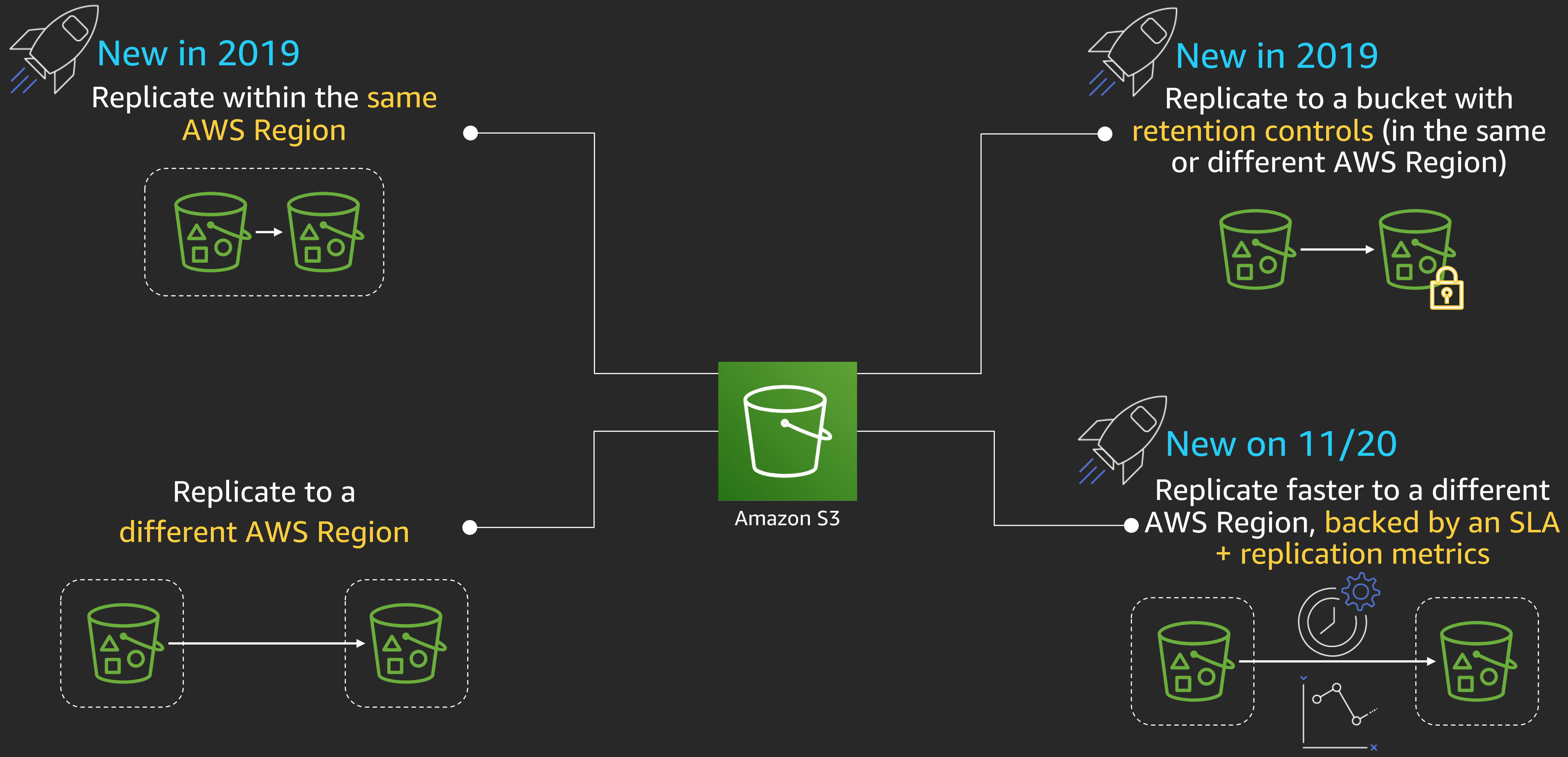


Amazon S3 Glacier Deep Archive

Features

- Lifecycle management
- Direct PUT to Amazon S3 Glacier or Deep Archive
- Amazon S3 Object Lock (WORM storage)
- Amazon S3 Glacier Restore Speed Upgrade
- Amazon S3 Glacier Restore Notifications
- Replicate direct to Amazon S3 Glacier or Deep Archive

Amazon S3 replication



Set S3 lifecycle policy to tier and expire storage



Use S3 storage class analysis to identify storage age groups that are less frequently accessed

S3 lifecycle policy to tier to lower-cost storage classes and expire objects. Fine-tune analysis by bucket, prefix, or object tag

For example, transition objects older than 180 days to S3 Glacier, and objects older than 365 days to S3 Glacier Deep Archive

You need the right tool to move data

So you can innovate anywhere

On-premises

File, block, tape

- AWS Storage Gateway

Online file and object transfer

- Rapid transfers: AWS DataSync
- File exchanges: AWS Transfer for SFTP
- Long-distance uploads and downloads: S3 Transfer Acceleration

Dedicated network

- AWS Direct Connect

Offline transfer

Bulk files, objects, HDFS, databases

- AWS Snowball Edge
- AWS Snowball
- AWS Snowmobile

AWS Partner Network Storage Technology Partners

CLOUDBERRY LAB IS NOW
MSP360

COHESITY

COMMVAULT

druva

IRON MOUNTAIN

TapeArk

VERITAS

Rock & Roll Hall of Fame

Heidi Quicksilver
Sr. Director Digital Systems and Strategy
Rock & Roll Hall of Fame

AWS re:Invent

aws



ROCK & ROLL
HALL OF FAME

The museum documents the history of rock music and the artists, producers, engineers, and other notable figures who have influenced its development.

Services

Amazon S3

Amazon S3 Glacier Deep Archive

AWS Snowball

Digital preservation with Amazon S3 Glacier Deep Archive

Opportunity

- New strategic plan to build RRHOF of the future, making the 21st century Rock Hall accessible worldwide and providing critical preservation of one of a kind digital media

Solution

- Partnered with AWS and Tape Ark to migrate all media into Amazon S3 via AWS Snowball, and set lifecycle policies in S3 to move media into Amazon S3 Glacier Deep Archive

Benefits

- Cost-effective storage to preserve estimated 350 TBs of data or approximately 2,800 hours of video material
- Recovered original preservation files through tape backups

Best practices

Decide on your data protection and retention strategy

Backup and Recovery

Accidental deletion
Existing backup mechanisms
Backup data retention

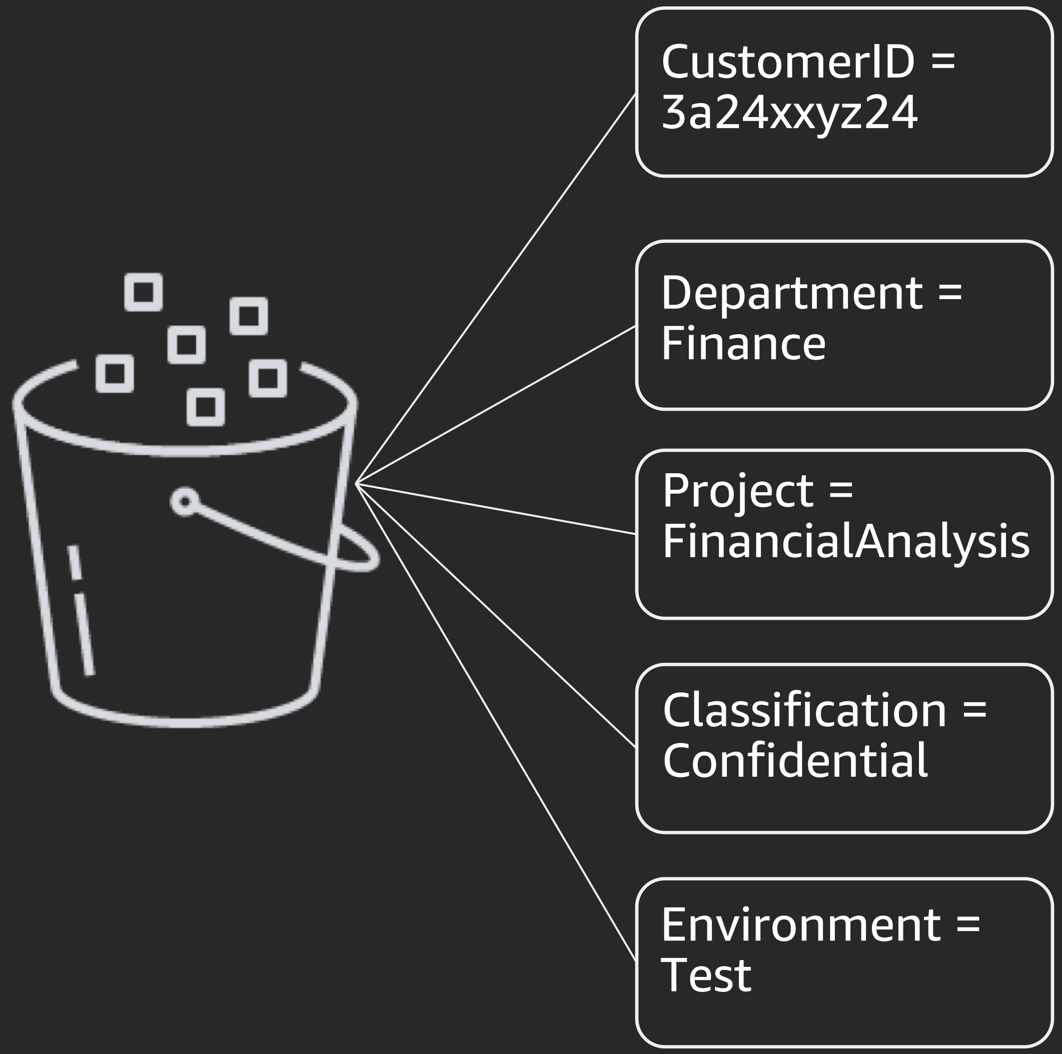
Data Retention

Data value
Access time
Compliance

Disaster Recovery

Recovery Point Objective
Recovery Time Objective
Geographic redundancy

Organize data with object tags and prefixes



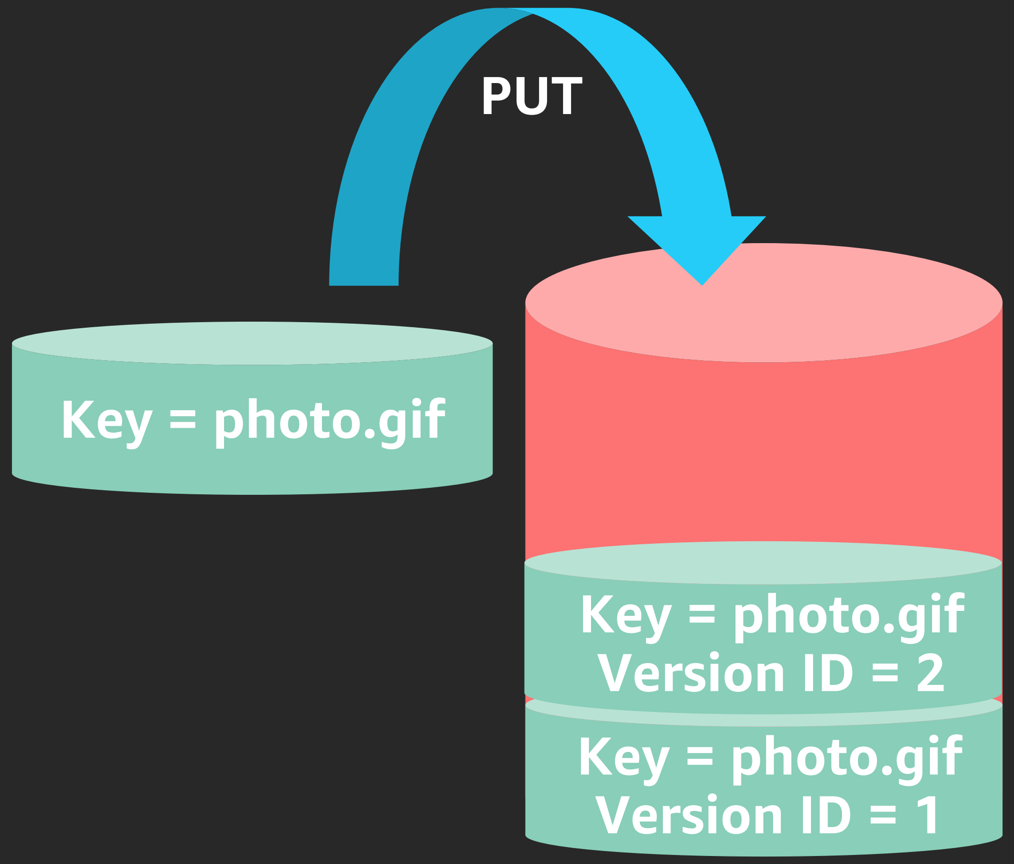
A single bucket can contain objects stored across any storage class

Control access, analyze usage, manage lifecycle policies, and replicate objects

Up to 10 mutable metadata tags (key value pair) per object

Tag objects when created, later, or both

Choose your versioning strategy



Create a new version with every upload. Previous versions are retained, not overwritten

Protect from unintended user deletes by enabling Multi-Factor-Authentication (MFA) on buckets

Making delete requests without a version ID removes access to objects but keeps the data

Manage previous versions with lifecycle. Transition or expire objects a specified number of days after they are no longer the current version

Need Immutability? – Amazon S3 Object Lock



Immutable S3 Objects

- Write Once Read Many (WORM) Protection for S3 Objects
- Object or bucket control of WORM & retention attributes

Retention Management Controls

- Define retention periods in your app or with bucket-level defaults
- Objects locked for the duration of the retention period
- Includes support for adding legal holds

Data Protection and Compliance

- Assessed for use in SEC 17a-4, CFTC, and FINRA environments
- Extra protection against accidental or malicious delete
- Audit retention date, mode, and legal hold status in S3 Inventory

Amazon S3 Object Lock – the basics

Ingest — Add retention controls to the object on PUT
… or use bucket defaults to set retention on all new objects

Lock — WORM enforced based on Retain Until Date
Versioning protects against overwrite, deletes are blocked

Retain — Use APIs to extend retention, or for legal holds

Audit — Review and audit retention information and object-level events

Accessing data in S3 Glacier Deep Archive

Restore before GET

- Issue a GET on an object in S3 Glacier and you’ll get a 408
- First issue a restore request for the object (select speed, duration of restore)
- S3 restores a copy of the object; optional notification issued
- Issue a GET request for the object (millisecond access)

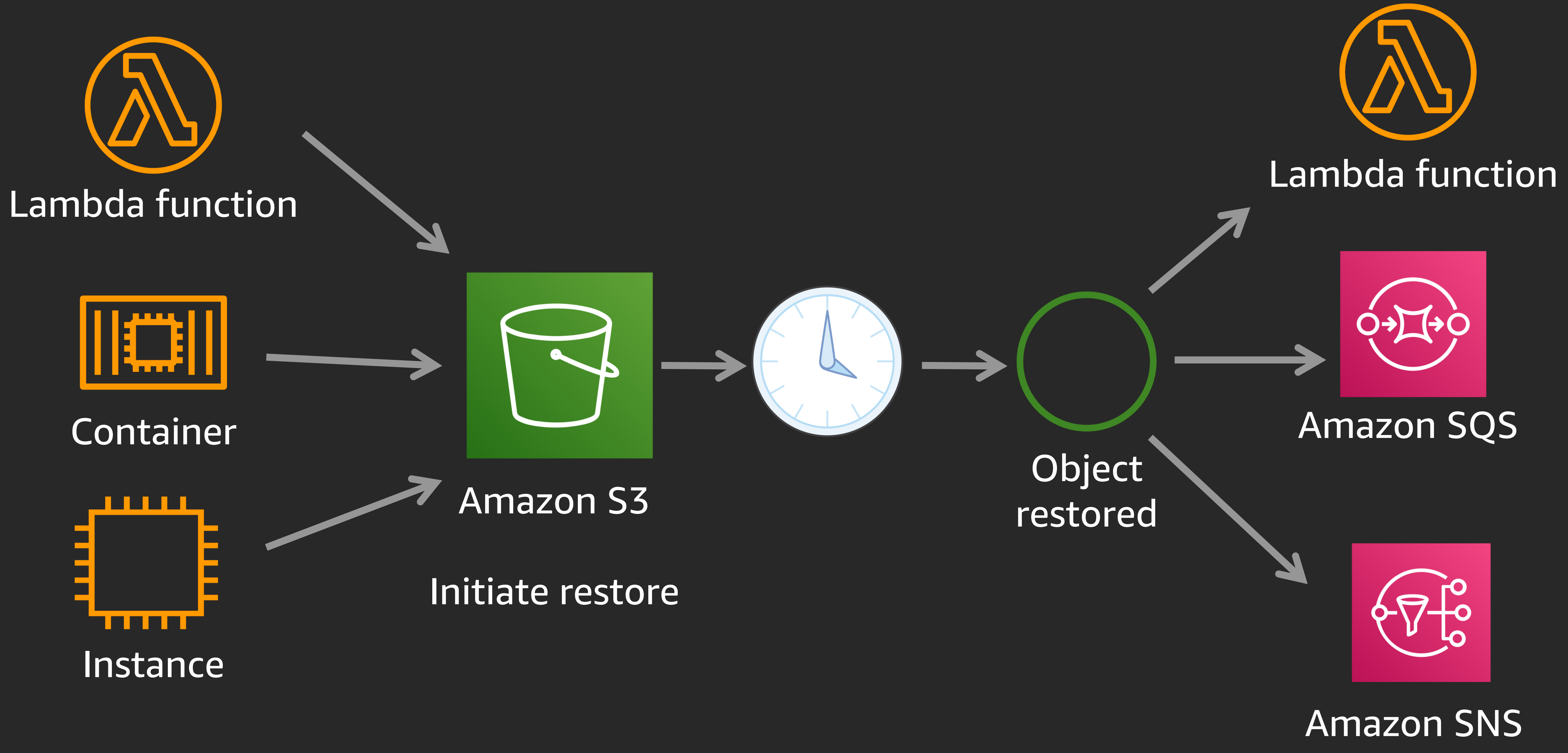
Use restore notifications to drive workflows through Amazon SNS, Amazon SQS, or AWS Lambda functions

```
$ aws s3api restore-object --bucket mybucket --key
dir1/example.obj --restore-request
'{"Days":25,"GlacierJobParameters":{"Tier":"Standard"}}'
```

Multiple restore speeds

	S3 Glacier	S3 Glacier Deep Archive
Expedited	Typically 1-5 minutes	Not available
Standard	Typically 3-5 hours	Typically within 12 hours
Bulk	Typically 5-12 hours	Typically within 48 hours

S3 Glacier restore – architecture pattern



Manage billions of objects with S3 Batch Operations



Choose Objects

- S3 inventory report
- CSV List

Use Case:

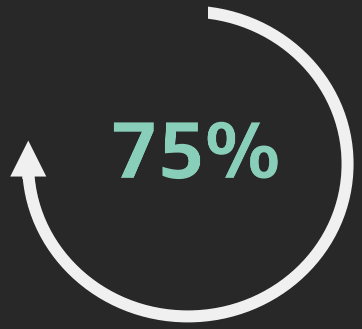
Create **S3 inventory report of objects to be restored**



Choose Operation

- Copy; restore from S3 Glacier; put access control list (ACL); replace object tag sets; run Lambda functions

Run **S3 Batch Operations** to restore objects



View Progress

- Object level progress
- Job notifications, automatic retries
- Completion report, auditing

Trigger Lambda functions with restore notifications

Cost considerations

S3 pricing dimensions



Ingest
(per-object)



Storage



Access
(per-object and per-GB)

Compare to on-premises tape libraries

Hosting (space, power, cooling, network)

Capital: servers, libraries, tapes

Tape generation migrations

Off-site storage & transportation

Administration & operations

Opportunity cost

Example workflow



50,000 objects



250 TB



0.5% accessed monthly

Long-term retention with Standard retrieval costs

		S3 Glacier	S3 Glacier Deep Archive
Ingest	50,000	$2.50 (@ $0.05/1K)	$2.50 (@ $0.05/1K)
Store	250 TB/M	$12,288 (@ $0.004/GB)	$3,041.28 (@ $0.00099/GB)
Access Bytes	1.25 TB/M (0.5%)	$153.6 (@ $0.01/GB)	$307.2 (@ $0.02/GB)
Access Objects	250 /M (0.5%)	$0.15 (@ $0.05/1K)	$0.30 (@ $0.10/1K)
	Annual cost	**$12,444.25**	**$3,351.28**

Long-term retention with Bulk retrieval costs

		S3 Glacier	S3 Glacier Deep Archive
Ingest	50,000	$2.50 (@ $0.05/1K)	$2.50 (@ $0.05/1K)
Store	250 TB/M	$12,288 (@ $0.004/GB)	$3,041.28 (@ $0.00099/GB)
Access Bytes	1.25 TB/M (0.5%)	$38.4 (@ $0.0025/GB)	$38.4 (@ $0.0025/GB)
Access Objects	250 /M (0.5%)	$0.075 (@ $0.025/1K)	$0.075 (@ $0.025/1K)
	Annual cost	**$12,328.97**	**$3,082.25**

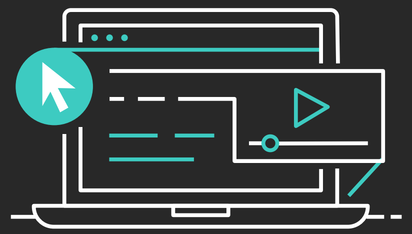
Putting it all together

- Understand your application requirements
- Understand your archival requirements
- Find the right mechanism to move data into S3
- Use tags and prefixes to organize and manage your data
- Understand the economics of archiving you data
- Archive to S3 Glacier Deep Archive for the lowest storage price

Questions?

Learn storage with AWS Training and Certification

Resources created by the experts at AWS to help you build cloud storage skills



45+ free digital courses cover topics related to cloud storage, including:

- Amazon S3
- AWS Storage Gateway
- Amazon S3 Glacier
- Amazon Elastic File System (Amazon EFS)
- Amazon Elastic Block Store (Amazon EBS)



Classroom offerings, like Architecting on AWS, feature AWS expert instructors and hands-on activities

Visit aws.amazon.com/training/path-storage/

aws training and certification

Thank you!

AWS re:Invent

Please complete the session survey in the mobile app.